Scalable Source Localization with Multichannel Alpha-Stable Distributions

Mathieu Fontaine, Charles Vanwynsberghe, Antoine Liutkus, Roland Badeau

SCALABLE SOURCE LOCALIZATION WITH MULTICHANNEL α-STABLE DISTRIBUTIONS

Mathieu Fontaine,[1] *Charles Vanwynsberghe,*[2] *Antoine Liutkus,*[1] *Roland Badeau*[3]

[1]Inria, speech processing team, Villers-lès-Nancy, France.
[2] Institut Jean le Rond d'Alembert, Saint-Cyr l'École, France.
[3]LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France.

ABSTRACT

In this paper, we focus on the problem of sound source localization and we propose a technique that exploits the known and arbitrary geometry of the microphone array. While most probabilistic techniques presented in the past rely on Gaussian models, we go further in this direction and detail a method for source localization that is based on the recently proposed α-stable harmonizable processes. They include Cauchy and Gaussian as special cases and their remarkable feature is to allow a simple modeling of impulsive and real world sounds with few parameters. The approach we present builds on the classical convolutive mixing model and has the particularities of requiring going through the data only once, to also work in the underdetermined case of more sources than microphones and to allow massively parallelizable implementations operating in the time-frequency domain. We show that the method yields interesting performance for acoustic imaging in realistic simulations.

***Index Terms*—source localization, acoustic modeling, α-stable random variables, spectral measure, sketching**

I. INTRODUCTION

Over the past few years, many studies have been conducted on the localization of audio objects, called *sources,* in an acoustic environment. The purpose of this trend of research may for instance be to recover speech from a noisy [1], [2], [3], [4] or reverberant [5], [2] environment, while knowing the geometry of the microphone array. In the scientific community, the incoming direction of propagating waves originating from a source is called a *direction of arrival* (DoA). Its knowledge is useful in a wide range of applications: sonar [6], [7], wireless communications [8], or medical imaging [9], [10] to cite a few. For the estimation of the DoAs, an efficient approach is to consider several candidate directions and the corresponding wave propagation as given by the physical model. In practice, this approach makes it possible to build the propagation operator from all candidate DoAs through the so-called *mixing matrix*, encapsulating the physical model. Several methods were proposed for estimating the DoAs. Among those, we may mention the celebrated MUSIC algorithm [11], which assumes that the number of sources is equal to the rank of the signal subspace and uses its orthogonal subspace called *noise subspace* for localization. The ESPRIT algorithm [12] is similar to MUSIC but exploits further rotational invariance properties for the signal subspace. Both algorithms achieve good localization, even though ESPRIT proved to be more robust against noise than MUSIC [13]. Another path of research in the case of audio localization is based on the Local Gaussian Model introduced in [14]. In this model, the dependences between the observations at different microphones are not modeled deterministically using classical features such as inter-channel time differences, but rather directly as frequency-dependent correlations embedded in the so-called *spatial covariance matrices.* In this context, the spatial covariance matrices corresponding to each DoA may be learned beforehand and used afterwards for localization as in [15].

The aforementioned methods for localization all exploit second-order statistics for the sources or observation signals, which are commonly modeled as wide-sense stationary (WSS) processes. From this perspective, it is common to pick a Gaussian process model for the signals [16], to be used along with the assumption of independent sources. Model parameters are then usually estimated by means of an expectation–maximization (EM) algorithm.

However, several studies pinned down the empirical fact that the marginal distributions of the signals of interest work poorly in accordance to the Gaussian model. In other terms, the histogram of the signals of interest is rarely correctly modeled by a Gaussian or a mixture of just a few Gaussian distributions. A classical option to address this problem is to adopt an heteroscedastic Gaussian model, meaning that we adopt varying mean and covariance parameters. In audio, this is notably the case of the Local Gaussian Model (LGM [14]), which operates on the Short Term Fourier Transform (STFT) and models all its entries as independent and centered Gaussian with different variances. As shown, e.g. in [17], this is equivalent to assuming the processes to be locally WSS. This approach yields good performance at the cost of requiring much more parameters to estimate than a simpler *marginal model*. That said, even when sticking to this *local stationarity* idea, the Gaussian model was recently outperformed in the context of single channel enhancement by its heavy-tailed α-stable generalization [18], coined in [19] as the α-harmonizable model. This framework encompasses the Cauchy ($\alpha = 1$) and Gaussian ($\alpha = 2$) special cases and leads to better filtering performance for the same computational complexity.

Picking an α-stable model for modeling time series has the advantage of better accounting for the impulsivity and noisiness of real-world signals, and the particular α-harmonizable case further permits convenient Time-Frequency (TF) processing. Still, the question of how to handle multichannel signals in this context remains unclear. Indeed, α-stable distributions have the noticeable disadvantage of not having finite second-order moments, prohibiting algorithms such as MUSIC or ESPRIT. Some research around this powerful framework has focused on beamforming under α-stable noise [20], but this does not translate to the case of α-stable target sources. More recent studies such as [21] considered Independent Component Analysis (ICA) for α-stable sources and successfully applied the method for the processing of heavy-tailed biomedical data in [22]. This is relevant in our setting because ICA and localization are related problems [23], [24] in the sense that estimating a mixing matrix is related to estimating DoAs. However, the approach in [21] suffers from several intrinsic limitations. Firstly, it considers sources that are independent and identically

This work was partly supported by the research programmes KAMoulox (ANR-15-CE38-0003-01) and EDiSon3D (ANR-13-CORD-0008-01) funded by ANR, the French State agency for research.

distributed (i.i.d.) in the time domain, which is a poor assumption in audio. Secondly, it relies on instantaneous mixing scenarios, for which the mixing matrix boils down to just scalar gains applied to the sources. While this is sufficient for electric or magnetic wavelengths, it is not realistic in an acoustic setting where the time taken by waves propagation is not negligible. Finally, the approach badly scales to massively multichannel arrays because it requires uniform gridding of the K-dimensional sphere where K is the number of microphones. This is prohibitive in the case of a few dozens or hundreds of microphones.

In this paper, we build on previous work on α-stable multivariate data [21], [25] but we extend the approaches to deal with those aforementioned drawbacks while providing exact inference. Firstly, we pick the α-harmonizable processes instead of i.i.d. α-stable sources. This has the advantage of accounting for realistic locally stationary impulsive signals. Secondly, we pick a marginal model for the sources in the Time-Frequency domain, meaning that each source is parameterized by only a very limited number of parameters, in contrast to the LGM model. The resulting localization method is not likely to be trapped in local minima for parameters estimation. Actually, the corresponding optimization problem we present is convex. Thirdly, we exploit the knowledge of the geometry of the microphone array along with physical assumptions about sound propagation to allow for exact inference even in massively multichannel arrays. This comes with a realistic modeling of convolutive mixing scenarios. Very interestingly, the resulting algorithm is finally of linear complexity: the data needs only to be processed once through a very simple *sketching* operation [26] to yield a very small representation from which the DoAs are estimated.

II. NOTATION, MODEL AND THEORY

II-A. Notation and acoustic model

We consider L possible positions in the 3-dimensional space. We then assume that each of them is the spatial location of a source s_l. Whenever no real source is actually present at that location, s_l should simply be estimated as 0. Our objective in this study becomes to estimate the magnitude of all s_l, removing the need of knowing the actual number of active sources. Now, picking an STFT representation with F frequency bands and T time frames, the sources are described by a tensor of size $F \times T \times L$, with complex entries. Hence, $s_l\left(f,t\right)$ is the complex spectrum of the l^{th}source at TF-bin (f,t). Likewise, x is the $F \times T \times K$ tensor gathering the STFTs of the signals received by K microphones. $\boldsymbol{x}\left(f,t\right)$ denotes the $K \times 1$ complex vector grouping the STFTs of all received signals at TF bin (f,t). The positions of the K microphones in $\mathbb{R}^3$ are assumed known. Classical acoustic propagation theory leads us to take the mixtures as a linear combination of the sources weighted by the complex $K \times 1$ steering vectors $\boldsymbol{A}_l\left(f\right)$, also called mixing filters [27]:

$$\forall\left(f,t\right),\ \boldsymbol{x}\left(f,t\right) \simeq \sum_{l=1}^{L} \boldsymbol{A}_l\left(f\right) s_l\left(f,t\right). \tag{1}$$

In the context of the classical near field region assumption [28], the steering vectors are for instance given by:

$$\forall l,\ f\ ,k,\ \left[\boldsymbol{A}_l\left(f\right)\right]_k = \frac{1}{r_{kl}} \exp\left(-i\frac{\omega_f r_{kl}}{c_0}\right), \tag{2}$$

where $r_{k,l}$ is the euclidean distance between the k^{th}microphone and the l^{th} source, ω_f is the angular frequency in frequency band f and c_0 is the sound speed in the air. In the general case, the steering vectors are provided by any relevant acoustical modeling of the room or even by actual measurements of the Room Impulse Responses (RIR). Given this acoustic model, we now turn to the isotropic α-stable model that we pick for the source signals s_l.

II-B. Isotropic α-stable distributions and source model

Let v be a complex random variable and φ_v its characteristic function (chf.). We say that v follows a symmetric isotropic α-stable distribution with $\alpha \in (0,2]$, denoted $S\alpha S_c$, if and only if φ_v has the following form:

$$\forall\theta \in \mathbb{C},\ \varphi_v\left(\theta\right) \triangleq \mathbb{E}\left(\exp\left(i\Re\left(\theta^{\star}v\right)\right)\right) = \exp\left(-\Upsilon\left|\theta\right|^{\alpha}\right), \tag{3}$$

where $\triangleq$ denotes a definition, $|.|$ denotes the modulus, $\cdot^{\star}$ denotes complex conjugation, $\Re\left(z\right)$ is the real part of $z \in \mathbb{C}$ and $\Upsilon \geq 0$ is a *scale parameter* that basically characterizes the amplitude of $v \sim S\alpha S_c(\Upsilon)$. This is a useful distribution for modeling impulsive data in signal processing [29], [30], [31]: the closer the *characteristic exponent* α is to 0, the heavier the tail is.

In this paper, we assume that each source s_l is an α-harmonizable process [19]. This boils down to taking the random variables $s_l\left(f,t\right)$ for all l, f, t as mutually independent, and distributed according to:

$$s_l\left(f,t\right) \sim S\alpha S_c\left(\Upsilon_l\right). \tag{4}$$

The scale parameters Υ_l can be viewed as the amplitudes of the latent sources and are equal to 0 when there is no source in the direction considered. As can be seen, the scale parameters Υ_l of the source signals are here taken as independent of both time and frequency. This strongly contrasts with the LGM model [14] for which they are time and frequency dependent. The fact is that picking a heavy-tail distribution such as $S\alpha S_c$ precisely allows for such a simple marginal model. Let $\boldsymbol{\Upsilon} \triangleq \left[\Upsilon_1, \ldots, \Upsilon_L\right]^{\top}$ be the main $L \times 1$ quantity of interest to estimate, coined in [32] as the *discrete spatial measure* (DSM).

II-C. Spatial measure for multichannel α-stable convolutive mixtures

According to (1) and (4), all vectors $\boldsymbol{x}\left(f,1\right), \ldots, \boldsymbol{x}\left(f,T\right)$ share the same distribution for a given frequency f. Moreover, it can be shown that they are symmetric $\alpha-$stable random vectors but they are not isotropic. Let φ_f be the chf. of the distribution of $\boldsymbol{x}\left(f,t\right)$ for any t:

$$\forall\boldsymbol{\theta} \in \mathbb{C}^K,\ \varphi_f\left(\boldsymbol{\theta}\right) \triangleq \mathbb{E}\left(\exp\left(i\Re\left\langle\boldsymbol{\theta}, \boldsymbol{x}\left(f,t\right)\right\rangle\right)\right). \tag{5}$$

where $\langle.,.\rangle$ is the inner product on $\mathbb{C}^K$. It can be uniquely expressed for $\alpha \in (0,2)$ as [18, p58]:

$$\forall\boldsymbol{\theta} \in \mathbb{C}^K,\ \varphi_f\left(\boldsymbol{\theta}\right) = \exp\left(-\int_{\boldsymbol{a} \in \mathscr{S}_{\mathbb{C}}^K} \left|\left\langle\boldsymbol{\theta}, \boldsymbol{a}\right\rangle\right|^{\alpha} d\Gamma_f\left(\boldsymbol{a}\right)\right), \tag{6}$$

where $\mathscr{S}_{\mathbb{C}}^K$ is the complex K-dimensional sphere $\mathscr{S}_{\mathbb{C}}^K \triangleq \left\{\left(z_1, \ldots, z_K\right) \in \mathbb{C}^K;\ \sum_{k=1}^K \left|z_k\right|^2 = 1\right\}$, and Γ_f is called the *spectral measure*, and is symmetric in the sense that for any continuous function φ defined on $\mathscr{S}_{\mathbb{C}}^K$ and for any $z \in \mathscr{S}_{\mathbb{C}}^1$, we have $\int_{\boldsymbol{a} \in \mathscr{S}_{\mathbb{C}}^K} \varphi(\boldsymbol{a}) d\Gamma_f\left(z\boldsymbol{a}\right) = \int_{\boldsymbol{a} \in \mathscr{S}_{\mathbb{C}}^K} \varphi(\boldsymbol{a}) d\Gamma_f\left(\boldsymbol{a}\right)$. We now introduce the quantity:

$$I_f\left(\boldsymbol{\theta}\right) \triangleq -\ln\left(\varphi_f\left(\boldsymbol{\theta}\right)\right), \tag{7}$$

called the Levy exponent [33] of the distribution of $\boldsymbol{x}\left(f,t\right)$. Combining (6) and (7), we get:

$$I_f\left(\boldsymbol{\theta}\right) = \int_{\boldsymbol{a} \in \mathscr{S}_{\mathbb{C}}^K} \left|\left\langle\boldsymbol{\theta}, \boldsymbol{a}\right\rangle\right|^{\alpha} d\Gamma_f\left(\boldsymbol{a}\right). \tag{8}$$

In this paper, we will exploit the Levy exponent for model estimation. To do this, it is desirable for computational reasons to approximate the integral in (8) as a finite sum. In practice, the sampling of a high dimensional sphere such as $\mathscr{S}_{\mathbb{C}}^K$ would raise the tricky issue of the *curse of dimensionality*. This is basically what

is done in [21] in the real case. To avoid it, the particular approach we propose in this paper is to exploit the additional information provided by the physical model (2).

Proposition 1. *In our case (1), $d\Gamma_f(\boldsymbol{a})$ is a sum of point masses:*

$$d\Gamma_f = \sum_{l=1}^{L} \Upsilon_l \|\boldsymbol{A}_l(f)\|^\alpha \delta_{\boldsymbol{a}_l(f)}, \tag{9}$$

where $\|.\|$ stands for the Hermitian norm and $\delta_{\boldsymbol{a}_l(f)}$ denotes the Dirac measure centered at $\boldsymbol{a}_l(f) \triangleq \boldsymbol{A}_l(f)/\|\boldsymbol{A}_l(f)\|$ and associated[1] to the equivalence relation $\boldsymbol{a} \sim \boldsymbol{b} \Leftrightarrow \exists z \in \mathscr{S}^1_{\mathbb{C}},\ \boldsymbol{a} = z\boldsymbol{b}$.

Proof: In short, we generalize [18, p70] to the complex case by substituting (1), (3) and (4) into (5) and by considering the independence of the sources, which yields:

$$\begin{aligned} \forall \boldsymbol{\theta} \in \mathbb{C}^K,\ I_f(\boldsymbol{\theta}) &= -\ln\left(\textstyle\prod_{l=1}^{L} \varphi_{s_l(f,t)}\left(\langle \boldsymbol{A}_l(f), \boldsymbol{\theta}\rangle\right)\right) \\ &= \sum_{l=1}^{L} \Upsilon_l \left|\langle \boldsymbol{\theta}, \boldsymbol{A}_l(f)\rangle\right|^\alpha. \end{aligned} \tag{10}$$

Since the same expression of $\varphi_f(\boldsymbol{\theta})$ can be obtained by substituting (9) into (6), and since the spectral measure Γ_f introduced in (6) is unique, (10) proves (9). ∎

In this study, for each frequency f, we estimate $\boldsymbol{\Upsilon}$ by exploiting relation (10) for L values $\boldsymbol{\theta}_1(f), \ldots, \boldsymbol{\theta}_L(f)$ of its operand. Assuming we have chosen those $\boldsymbol{\theta}_l(f)$, let:

$$\boldsymbol{I}_f \triangleq \left[I_f(\boldsymbol{\theta}_1(f)), \ldots, I_f(\boldsymbol{\theta}_L(f))\right]^\top \tag{11}$$

be the (nonnegative) $L \times 1$ vector gathering the Levy exponents of the mixture at frequency f for these $\boldsymbol{\theta}_l(f)$. Defining now the matrix $\boldsymbol{\Psi}_f$ of size $L \times L$ by:

$$[\boldsymbol{\Psi}_f]_{l,l'} \triangleq \left|\langle \boldsymbol{\theta}_l(f), \boldsymbol{A}_{l'}(f)\rangle\right|^\alpha, \tag{12}$$

equation (10) leads to $\boldsymbol{I}_f = \boldsymbol{\Psi}_f \boldsymbol{\Upsilon}$. Since this equality holds for all f, we can concatenate all $\boldsymbol{I}_f$ into the $FL \times 1$ vector $\boldsymbol{I}$ and all $\boldsymbol{\Psi}_f$ into the known $FL \times L$ matrix $\boldsymbol{\Psi}$. We obtain:

$$\boldsymbol{I} = \boldsymbol{\Psi}\boldsymbol{\Upsilon}, \tag{13}$$

which is our starting point for estimating $\boldsymbol{\Upsilon}$. In practice, we pick $\boldsymbol{\theta}_l(f) = \boldsymbol{A}_l(f)$, because this choice proved effective and leads to a symmetric matrix $\boldsymbol{\Psi}_f$. Note however that other options may be considered.

III. PARAMETER ESTIMATION

We now present a way to estimate $\boldsymbol{\Upsilon}$ using (13). This method is divided into two steps: firstly, we estimate $\boldsymbol{I}$ from the data and secondly, we estimate $\boldsymbol{\Upsilon}$ based on the knowledge of this estimate, the knowledge of $\boldsymbol{\Psi}$ and the equality $\boldsymbol{I} = \boldsymbol{\Psi}\boldsymbol{\Upsilon}$.

III-A. Estimation of $\boldsymbol{I}$: empirical levy exponent

The classical empirical characteristic function (ECF) of the observations $\boldsymbol{x}(f,1), \ldots, \boldsymbol{x}(f,T)$, denoted $\widetilde{\varphi}_f$, is defined as [34]:

$$\forall \boldsymbol{\theta} \in \mathbb{C}^K,\ \widetilde{\varphi}_f(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \exp\left(i\Re\left\langle \boldsymbol{\theta}, \boldsymbol{x}(f,t)\right\rangle\right). \tag{14}$$

It would be logical to apply (14) to estimate $\boldsymbol{I}_f$ as $\tilde{\boldsymbol{I}}_f = -\ln\left(\widetilde{\varphi}_f(\boldsymbol{\theta})\right)$. However, taking the logarithm of (14) may raise an issue because $\tilde{\varphi}_f$ could be complex-valued numerically. Fortunately, in the symmetric α-stable case, we can address this issue in the following way:

Proposition 2. *In the $S\alpha S_c$ case, we can define an unbiased estimator of the chf. as:*

$$\forall \boldsymbol{\theta} \in \mathbb{C}^K,\ \widehat{\varphi}_f(\boldsymbol{\theta}) = \left|\frac{1}{T}\sum_{t=1}^{T} \exp\left(i\frac{\Re\left\langle \boldsymbol{\theta}, \boldsymbol{x}(f,t)\right\rangle}{2^{1/\alpha}}\right)\right|^2. \tag{15}$$

Proof: From the α-stability assumption, we have that the two sets of observations: $\{\boldsymbol{x}(f,t)\}_t$ and $\left\{\left(2^{-1/\alpha}\right)\left(\boldsymbol{x}(f,t_1) - \boldsymbol{x}(f,t_2)\right)\right\}_{t_1,t_2}$ share the same probability distribution. Replacing the first one by the second one in (14), and factorizing the result, we finally get (15). ∎

Since $\widehat{\varphi}_f(\boldsymbol{\theta})$ is nonnegative, we can define the *empirical Levy exponent*: $\widehat{\boldsymbol{I}}_f(\boldsymbol{\theta}) = -\ln\widehat{\varphi}_f(\boldsymbol{\theta})$.

The strategy we adopt is simply to compute the empirical Levy exponent $\hat{\boldsymbol{I}}_f(\boldsymbol{\theta})$ of the data for all $\boldsymbol{\theta}_l(f) = \boldsymbol{A}_l(f)$. As in section II-A, we put all those LF estimates together and note them $\widehat{\boldsymbol{I}}$. Following (13), they obey:

$$\widehat{\boldsymbol{I}} \approx \boldsymbol{\Psi}\boldsymbol{\Upsilon}. \tag{16}$$

It is fundamental to note here that the estimate $\widehat{\boldsymbol{I}}$ of the Levy exponent is obtained after only one pass over the data through a very simple procedure (15). Parameter estimation is then achieved from this compressed $LF \times 1$ vector $\widehat{\boldsymbol{I}}$ only, through $\widehat{\boldsymbol{I}} \approx \boldsymbol{\Psi}\widehat{\boldsymbol{\Upsilon}}$, and not using the much larger $F \times T \times K$ observation dataset $\boldsymbol{x}$. We see we have here an example of the *sketching* framework recently discussed in [26], that naturally arises when considering α-stable random variables. This fact means that the proposed estimation procedure is not only of linear complexity, it is massively parallelizable over both time and frequency.

III-B. Estimation of $\boldsymbol{\Upsilon}$ with nonnegative optimization

Our objective is the estimation of the nonnegative quantity $\boldsymbol{\Upsilon}$, so that (16) is verified. A natural idea for this purpose is to estimate it so as to minimize the discrepancies between the left and right-hand sides of (16), by picking:

$$\widehat{\boldsymbol{\Upsilon}} \leftarrow \arg\min_{\boldsymbol{\Upsilon} \geq 0} \sum_{f,l} d_\beta\left(\widehat{\boldsymbol{I}} \mid \boldsymbol{\Psi}\boldsymbol{\Upsilon}\right) \tag{17}$$

where d_β is any data-fit cost function such as the $\beta-$divergence [35]: $d_\beta(x|y) = \frac{1}{\beta(\beta-1)}(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1})\ \forall x > 0, y > 0, \beta \in \mathbb{R}\backslash\{0,1\}$. To solve this optimization problem, we note that all quantities involved in (17) are nonnegative, so that it is natural to adopt a now-classical multiplicative update strategy as in nonnegative matrix factorization (NMF, [36]) and to iterate the following update:

$$\widehat{\boldsymbol{\Upsilon}} \leftarrow \widehat{\boldsymbol{\Upsilon}} \cdot \frac{\boldsymbol{G}_-\left(\widehat{\boldsymbol{\Upsilon}}\right)}{\boldsymbol{G}_+\left(\widehat{\boldsymbol{\Upsilon}}\right)}, \tag{18}$$

where $a \cdot b$ and $\frac{a}{b}$ stand for element-wise multiplication and division of a and b, respectively. $\boldsymbol{G}_-\left(\widehat{\boldsymbol{\Upsilon}}\right)$ and $\boldsymbol{G}_+\left(\widehat{\boldsymbol{\Upsilon}}\right)$ are obtained by the following calculation:

$$\nabla_{\widehat{\Upsilon}} d_\beta\left(\widehat{\boldsymbol{I}} | \boldsymbol{\Psi}\widehat{\boldsymbol{\Upsilon}}\right) = \underbrace{\boldsymbol{\Psi}^\top\left(\left(\boldsymbol{\Psi}\widehat{\boldsymbol{\Upsilon}}\right)^{\beta-1}\right)}_{\boldsymbol{G}_+\left(\widehat{\boldsymbol{\Upsilon}}\right)} - \underbrace{\boldsymbol{\Psi}^\top\left(\left(\boldsymbol{\Psi}\widehat{\boldsymbol{\Upsilon}}\right)^{\beta-2} \cdot \widehat{\boldsymbol{I}}\right)}_{\boldsymbol{G}_-\left(\widehat{\boldsymbol{\Upsilon}}\right)}. \tag{19}$$

[1] In other words, for any continuous function φ defined on $\mathscr{S}^K_{\mathbb{C}}$, $\int_{\boldsymbol{a} \in \mathscr{S}^K_{\mathbb{C}}} \varphi(\boldsymbol{a}) \delta_{\boldsymbol{a}_l(f)}(\boldsymbol{a}) = \int_{z \in \mathscr{S}^1_{\mathbb{C}}} \varphi(z\,\boldsymbol{a}_l(f)) \frac{dz}{2\pi}$.

The algorithm box below summarizes this approach.

Algorithm 1 Estimation of the spatial measure $\widehat{\boldsymbol{\Upsilon}}$

1) **Input**
   - Number L of possible positions.
   - Steering vectors $\boldsymbol{A}_l(f)$ as in (2) or directly using RIR.
   - STFTs $\boldsymbol{x}$ of the mixture, of size $F \times T \times K$.
   - Number of iterations.
   - Parameters α and divergence to be used.
2) **Sketching**:
   for all f, compute $\widehat{\boldsymbol{I}}_f\left(\boldsymbol{A}_l\left(f\right)\right) = -\ln \widehat{\varphi}_f\left(\boldsymbol{A}_l\left(f\right)\right)$ using (15)
3) **Parameter estimation**
- Compute $\boldsymbol{\Psi}_f$ according to (12) with $\boldsymbol{\theta}_l\left(f\right) = \boldsymbol{A}_l\left(f\right)$.
- Compute $\widehat{\boldsymbol{\Upsilon}}$ by iterating over (18).

IV. EVALUATION

In this section, we investigate the performance of the proposed method with room impulse responses RIRs[2], simulating a room of dimensions $5 \times 4 \times 3$ m with 0.4 s of reverberation time [37]. The sources are taken as voice samples from the CMU[3] dataset with sample rates of 16 kHz and distributed randomly on a 5×4 plane, 1.5 m above the floor. The candidate locations for analysis are positioned on a grid of 10 cm step size, corresponding to $L = 2091$, and 5 cm shifted with respect to the true sources' locations to ensure realistic experiments. Then, the observations $\boldsymbol{x}\left(f,t\right)$ are obtained by propagating the sources the RIR (2). The excerpts considered last 15 seconds ($T = 391$) and we use only the $F = 128$ frequency bands corresponding to the interval $[1000,\ 3000]$ Hz.

We pick $\alpha = 1.2$ after [19], and we set the beta divergence parameter to $\beta = 0$ (Itakura-Saito divergence). We run 70 iterations of the NMF algorithm in step 3) of Algorithm 1.

The baseline method we use here for comparison is the *steered response power* (SRP, [38]). It consists in backpropagating the observations towards the sources locations using the steering vectors $\boldsymbol{A}_l\left(f\right)$. Basically, it computes simple inner products between $\boldsymbol{x}\left(f,t\right)$ and $\boldsymbol{A}_l\left(f\right)$. We define the SRP measure $\widehat{\boldsymbol{B}} \triangleq \left[\widehat{B}_1, \ldots, \widehat{B}_L\right]^\top$ as:

$$\forall l,\ \widehat{B}_l = \frac{1}{FT} \sum_{f,t} \left|\boldsymbol{A}_l^\star\left(f\right) \boldsymbol{x}\left(f,t\right)\right|. \tag{20}$$

To compare the performance of both approaches, we first define the ground truth as the support of the true DSM $\boldsymbol{\Upsilon}$, that is: 1 in the active directions and 0 otherwise, followed by a smoothing using a Gaussian kernel with a 10cm length-scale. We compute a total number of 200 simulations, i.e. 20 independent trials for each case of $J = 1, \ldots, 4$ sources and $K = 3, 4, 10, 20, 30$ randomly positioned microphones.

We first note in Fig. 1 that the score obtained with SRP is much lower than the correlation obtained with the proposed method. After investigation, it turns out that this bad performance is mostly due to the fact that much of the energy of the SRP is spread out across many directions. We highlight that a large number of microphones delivers better results ($\simeq 65\%$ for $K = 30$ and $\simeq 21\%$ for $K = 3$).

[2]https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator

[3]Carnegie Mellon University dataset : http://www.festvox.org/cmu_faf/

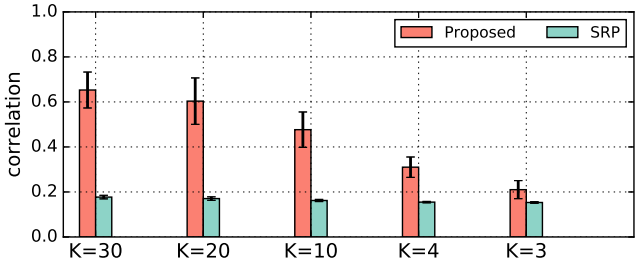

Fig. 1. Scores of correlations with ground truth. Black whiskers depicted the deviations.

Still, a very interesting feature of the proposed approach is that it yields a DoA heatmap (the DSM) that is intrinsically very sparse, contrarily to SRP, as can be seen with a typical example shown on Fig. 2.

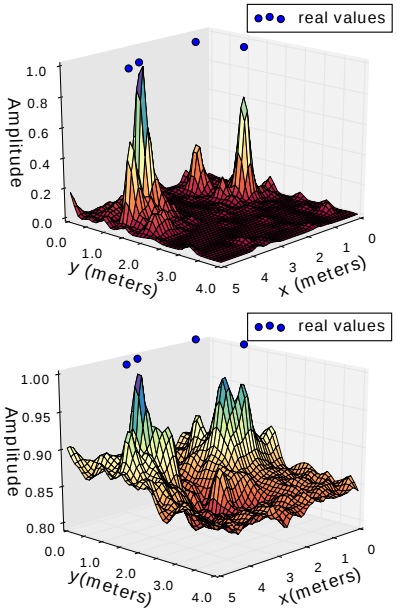

Fig. 2. Localization heatmap with the spatial measure $\widehat{\boldsymbol{\Upsilon}}$ (top) and the SRP $\widehat{\boldsymbol{B}}$ (bottom) for 30 microphones and 4 sources.

V. CONCLUSION

In this paper, we proposed an α-stable multivariate probabilistic model for multichannel audio recordings and we have shown that it leads to the concept of the discrete spatial measure (DSM), which is directly related to the directions of arrivals of the sources. Exploiting acoustics, we have proposed a method to estimate the DSM, that requires going through the data only once and that is of linear complexity, making it possible to process streams of massively multichannel audio. We have shown that the method considerably outperforms classical beamforming approaches in terms of acoustic imaging and showed its interest in blindly imaging virtual sources and thus the shape of the room through the exploitation of the echoes.

This study would now benefit from more challenging experiments where the method would be compared with more robust algorithms as CLEAN and RELAX. Another step into the α-stable theory would be to augment the model with elliptically contoured distributions, permitting the introduction of some uncertainty on the steering vectors.

VI. REFERENCES

[1] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, 1999.

[2] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Signal Process.*, vol. 34, no. 6, pp. 1391–1400, 1986.

[3] Y. Grenier, "A microphone array for car environments," *Speech Communication*, vol. 12, no. 1, pp. 25–39, 1993.

[4] C.E. Chen, F. Lorenzelli, R.E. Hudson, and K. Yao, "Stochastic maximum-likelihood DOA estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3038–3044, 2008.

[5] G. Huang, J. Chen, and J. Benesty, "Direction-of-arrival estimation of passive acoustic sources in reverberant environments based on the Householder transformation," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3053–3060, 2015.

[6] S. Haykin, J.H. Justice, N.L. Owsley, J.L. Yen, and A.C. Kak, *Array signal processing*, Prentice-Hall, Inc., Englewood Cliffs, NJ, USA, 1985.

[7] D. Rouseff and L.M. Zurk, "Striation-based beamforming for estimating the waveguide invariant with passive sonar," *The Journal of the Acoustical Society of America*, vol. 130, no. 2, pp. EL76–EL81, 2011.

[8] J. Mietzner, R. Schober, L. Lampe, W.H. Gerstacker, and P.A. Hoeher, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 2, pp. 87–105, 2009.

[9] H. Hasegawa and H. Kanai, "High-frame-rate echocardiography using diverging transmit beams and parallel receive beamforming," *Journal of medical ultrasonics*, vol. 38, no. 3, pp. 129–140, 2011.

[10] T. Hergum, T. Bjastad, and H. Torp, "Parallel beamforming using synthetic transmit beams [biomedical ultrasound imaging]," in *Proc. of IEEE Ultrasonics Symposium*. IEEE, 2004, vol. 2, pp. 1401–1404.

[11] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[12] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[13] Y. Hua and T.K. Sarkar, "On SVD for estimating generalized eigenvalues of singular matrix pencil in noise," in *Proc. of IEEE International Symposium on Circuits and Systems*. IEEE, 1991, pp. 2780–2783.

[14] N.Q.K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 18, no. 7, pp. 1830 –1840, sept. 2010.

[15] J. Nikunen and T. Virtanen, "Direction of arrival based spatial covariance model for blind sound source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 3, pp. 727–739, 2014.

[16] R.J. Kozick and B.M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 601–616, 2004.

[17] A. Liutkus, R. Badeau, and G. Richard, "Gaussian processes for underdetermined source separation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3155–3167, 2011.

[18] G. Samoradnitsky and M. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*, vol. 1, CRC Press, 1994.

[19] A. Liutkus and R. Badeau, "Generalized Wiener filtering with fractional power spectrograms," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 266–270.

[20] P. Tsakalides, *Array signal processing with alpha-stable distributions*, Ph.D. thesis, University of Southern California, Los Angeles, CA, USA, 1995.

[21] P. Kidmose, "Independent component analysis using the spectral measure for alpha-stable distributions," in *Proc. of IEEE-EURASIP 2001 Workshop on Nonlinear Signal and Image Processing*, 2001, vol. 400.

[22] V. Kiviniemi, J. Kantola, J.H. Jauhiainen, A. Hyvärinen, and O. Tervonen, "Independent component analysis of nondeterministic fMRI signal sources," *Neuroimage*, vol. 19, no. 2, pp. 253–260, 2003.

[23] H Sawada, R Mukai, S Araki, and S Makino, "Multiple source localization using independent component analysis," in *Antennas and Propagation Society International Symposium, 2005 IEEE*. IEEE, 2005, vol. 4, pp. 81–84.

[24] T Otsuka, K Ishiguro, H Sawada, and H G Okuno, "Bayesian unification of sound source localization and separation with permutation resolution.," in *AAAI*, 2012.

[25] D. Fitzgerald, A. Liutkus, and R. Badeau, "Projection-based demixing of spatial audio," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1560–1572, Sept. 2016.

[26] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2016, pp. 6190–6194.

[27] S Leglaive, R Badeau, and G Richard, "Multichannel audio source separation with probabilistic reverberation priors," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 12, pp. 2453–2465, Sept. 2016.

[28] H.L. Van Trees, *Detection, estimation, and modulation theory, optimum array processing*, John Wiley & Sons, 2004.

[29] P.G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, 1999.

[30] M. Sahmoudi, *Processus alpha-stables pour la séparation et l'estimation robustes des signaux non-gaussiens et/ou non-stationnaires*, Ph.D. thesis, Université Paris 11, Paris, France, 2004.

[31] A. Liutkus, T. Ulubanjo, E. Moore, and M. Ghovanloo, "Source separation for target enhancement of food intake acoustics from noisy recordings," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2015, pp. 1–5.

[32] M Fontaine, C Vanwynsberghe, A Liutkus, and R Badeau, "Sketching for nearfield acoustic imaging of heavy-tailed sources," in *13th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2017)*, 2017.

[33] M. Unser and P. Tafti, *An introduction to sparse stochastic processes*, Cambridge University Press, 2014.

[34] J.P. Nolan, A.K. Panorska, and J.H. McCulloch, "Estimation of stable spectral measures," *Mathematical and Computer Modelling*, vol. 34, no. 9, pp. 1113–1122, 2001.

[35] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[36] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[37] EAP Habets, "Room impulse response (rir) generator," 2008.

[38] D N Zotkin and R Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 499–508, 2004.